bio-protocol Preprint

Not Peer Reviewed
Ask a question online

Cluster diversity and isolation indices

DR Daniel A Russell GH Graham F Hatfull

Updated date: May 12, 2021

An abbreviated version of this protocol was published in eLIFE in Apr 2015

Whole genome comparison of a large collection of mycobacteriophages reveals a continuum of phage genetic diversity

DOI: 10.7554/eLife.06416

Detailed protocol

The only efficient way to do these comparisons is using some SQL and scripting. You'll need the Phamerator database underlying these calculations, which is available from http://databases.hatfull.org/Actino_Draft/. It's the "Actino_Draft.sql" file. You'll also need an SQL client on your computer to enable the running of queries on that database. The database is a MySQL database, so this should be the place to start: https://www.mysql.com/downloads/. We scripted in python, using a connector to access the database: https://dev.mysql.com/doc/connector-python/en/connector-python-example-connecting.html.
Bio-python is also valuable for these calculations: https://biopython.org/.
The formulas for the actual calculations are in the paper, so from there it's a matter of querying the database for the phams of interest, and doing the math to compare them. Since these sorts of scripts are heavily computer- and environment-dependent, it doesn't really make sense for us to share ours,

How to cite:（Readers should cite both the Bio-protocol preprint and the original research article where this protocol was used）

1. Russell, D. and Hatfull, G. (2021). Cluster diversity and isolation indices. Bio-protocol Preprint. bio-protocol.org/prep1079.
2. Pope, W. H., Bowman, C. A., Russell, D. A., Jacobs-Sera, D., Asai, D. J., Cresawn, S. G., Jacobs, W. R., Hendrix, R. W., Lawrence, J. G. and Hatfull, G. F.(2015). Whole genome comparison of a large collection of mycobacteriophages reveals a continuum of phage genetic diversity. eLIFE. DOI: 10.7554/eLife.06416

Copyright: Content may be subjected to copyright.